图书情报工作
第62卷 第4期 2018年2月

知识进化视角下科学文献传播网络演化与预测研究及应用*

丁玉飞 关鹏
南京理工大学经济管理学院 南京 210094

摘要：[目的/意义]分析与研究科学文献传播网络发展趋势，为揭示科研新成果、把握科技发展脉络以及发现学科领域内前沿发展动态提供参考借鉴。[方法/过程]阐述知识进化理论及其应用，探讨知识进化视角下的科学文献传播网络预测方法可行性，通过对科学文献传播网络的演化机制分析，构建科学文献传播网络预测模型，并以科学文献关键词网络进行实证分析。[结果/结论]通过科学文献数据实证分析表明，提出的知识进化视角下科学文献传播网络演化与趋势预测方法具有可行性和有效性，可以为科学文献传播网络的趋势发展分析提供新的研究思路。

关键词：知识进化 科学文献 传播网络 演化 预测
分类号：G250
DOI：10.13266/j.issn.0252－3116.2018.04.010

1 引言

随着科技创新环境和关联要素的日益多元化和动态复杂化，如何从海量科学文献中准确而快速地梳理相关领域研究的发展动态，挖掘出研究主题与演化趋势，找到科技创新突破口，成为科学发挥重要作用的基础。科学文献作为科技创新体系中重要的传播载体，是科学研究成果的主要表现形式，其中蕴含着巨大的社会价值和经济价值历来备受社会各界的重视和好评。科学文献旨在揭示科研新成果，体现学科领域内前沿的科技水平与发展动态[1]。近年来，由于科学技术的快速发展和数字出版行业及其相关技术的日渐成熟，科学文献的数量也与日俱增，从主题复杂多样、数量庞大的文献资源中获取学科领域的研究前沿和热点，把握科技发展的脉络，成为科研人员面临的共同难题[2]。

在科学文献传播过程中，科学文献依附的传播要素与其承载的知识内容单元，沿着多个知识节点交互传递与扩散，知识单元之间的相互关联与衔接关系作为纽带，由此构成了学科领域内的科学文献传播网络。从传播网络节点的组成内容分析，科学文献是由表达知识主体的外部单元和表达知识客体的内容单元构成的，其中，知识主体的外部单元如作者、机构、出版物来源、引文等，是知识传播的主体，具有掌控知识传播的主动权与选择权，控制与影响着知识进展与演化，但是相对于科学文献构成要素来讲属于外部单元。而知识客体的内容单元如关键词、主题、专业词汇、学科分类与全文等，是知识传播的客体，由知识主体驾驭的同时反作用于知识主体的行为，是知识进展与演化结果的反映，相对于科学文献构成要素来讲属于内容单元。从传播网络节点之间的关系与形成的不同结构来看，知识主体的外部单元与知识客体的内容单元的相互作用及其形成的不同关联，促进了科学文献的多元多向的传递与分享，由此带来了知识进展与演化的复杂性与多样性。所以，选择科学文献传播网络作为研究对象，研究知识主体的外部单元与知识客体的内容单元的关联及其作用，可以探究科学文献传播要素之间复杂演化的特点和规律，进而揭示知识进展的发展趋向，为科学技术决策提供依据与参考支撑。

目前，已有众多学者参与科学文献传播网络方面的研究，并形成了一定的研究基础。众多的研究发现，基于科学文献传播网络研究学科领域知识发展动态，可以突破单一要素研究的局限性，使研究结果更有效地呈现知识传播与分享的动态性与复杂性。同时，科

* 本文系国家自然科学基金项目"新研究领域科学文献传播网络生长及对传播效果影响研究"(编号:71373124)研究成果之一。
作者简介：丁玉飞(ORCID:0000－0002－0180－6546)，博士研究生，E-mail:dingyufei999@126.com；关鹏(ORCID:0000－0002－2308－3019)，博士研究生。
收稿日期:2017－10－17 修回日期:2017－12－01 本文起止页码:72－80 本文责任编辑:王善军

学文献传播网络作为科学知识的承载体,同样遵循着科学趋势变化是在先前研究的基础上衍生、拓展和创新而来的法则,这正是科学文献传播网络知识进化的体现。因此,笔者拟以知识进化理论为视角,分析科学文献传播网络的动力机制,探索其未来的动态发展趋势,为揭示科研新成果、挖掘科学领域研究热点提供研究思路。

2 知识进化及其主要研究内容

2.1 知识进化论的兴起

知识进化理论起源于达尔文进化论,即:生物之间存在着生存争斗,适应者生存下来,不适者则被淘汰,这就是自然选择的过程。而生物正是通过遗传、变异和自然选择,从低级到高级,从简单到复杂,种类由少到多地进化与发展[3]。在进化论思想的启发下,学者们开始探讨知识领域存在的生物进化的现象,并且在考察人类知识增长的过程中引入了进化的机制。如英国哲学家卡尔·波普尔提出科学知识的进化是一个类似于达尔文所说的自然选择的过程,即假说的自然选择,我们的知识总是由假说构成的,它通过在生存斗争中存在下来而显示了它的相对适应性,竞争消除了那些不适应的假说[4]。唐纳德·坎贝尔从生物进化的历程中得到启示,把知识进化分为了三个历程:①生成或创造新思想;②借助批评和检测淘汰那些不正确的思想;③传播那些尚未证明为不正确的思想[5]。从哲学的视角看待世间万物,它们都是处于不断的进化和演变当中的。因此,知识进化可以被认为是旧有的知识与新知识、旧思想与新思想不断更替的过程。

2.2 知识进化的机制与主要研究内容

依据知识进化理论,现有的新知识的产生和发展必须符合环境的需要。新知识的产生来自于对旧有知识的实践和检验,由此得到社会的认可。这一过程中,有些知识符合规律,而有些知识违背规律,经过长时间的自我检验和修正后,与客观事实相符的知识得以保留,反之则被淘汰。知识进化的过程往往也是同客观规律相适应的过程,主要包括知识基因、遗传机制、变异机制和自然选择机制等内容,下面对此进行探讨与分析。

(1)知识基因。由于知识进化的过程类似于生物的进化,也有知识基因的存在。从知识的构成发现,知识基因是由核心概念及核心概念之间的关系、独特的思维方式和学术规范所组成。知识基因是知识的内核,知识的进化方向本质上也是知识基因质量的跃迁过程[6]。随着科学技术的发展和知识领域的延伸,知识以知识基因为基本单位,通过遗传和变异等方式实现知识的进化。由此,刘植惠将知识基因定义为知识遗传与变异的最小功能单元[7]。他从知识可视化的角度提出知识元是知识领域的基本单位,一般是以词语、概念、术语来表征的信息内容,在一定条件下,某个关键的知识元可能扮演知识基因的角色,并决定着科学知识的进化过程。

(2)知识的遗传。知识的发展具有一定的延续性,任何知识的进化都是点滴的积累而形成的,从量变到质变的原理反映了知识累积的一个过程[8]。而知识的遗传推动了知识的不断增长。知识的遗传是指知识以知识基因的方式继承和传递下去,实现知识之间的延续性。有些知识符合规律,而有些知识违背规律,经过长时间的自我检验和修正后,与客观事实相符的知识得以保留,反之则被淘汰。知识的遗传过程则保留了适应的部分,去除掉不适应的部分,从而逐渐构建了知识体系。

(3)知识的变异。知识变异是指受到众多因素的影响,不同的知识基因片段重新组合在一起形成新的知识基因的过程[9]。这种变异的过程有可能发生在知识进化发展的各个阶段,知识的变异既是一个过程又是一个结果,不同的知识基因片段彼此之间发生接触,通过知识基因之间的碰撞就会有变异的产生。但是,知识的变异与知识的创新有着本质的区别,知识的变异只有被认可后并进行扩散,才能上升到知识创新的层面。

(4)知识的自然选择。由于不同知识主体对于科学知识的认知存在差异,通过社会选择的方式来确定那些知识是否被保留或淘汰,称之为知识的自然选择[10]。与生物进化的选择过程相比,知识的自然选择更复杂。生物的自然选择具有随机性,而知识的选择是带有主观目的性。知识主体提出的科学知识只有经过知识群体的认可,才能被传播、分享与利用,这个选择的过程相对复杂,涉及的因素众多,如社会政治、经济和文化环境等多重因素都对知识的选择有着重大的影响。正是由于知识存在着这样的自然选择过程,才促使人类在探索客观世界的过程中不断接近真理。

目前,知识进化的研究正逐渐得到社会各界的关注,获得许多重要的研究成果,并应用于众多领域。从知识进化理论的应用现状来看知识进化的研究与应用具有一定的理论意义与实践意义的。科学文献是承载知识的载体,体现了学科领域的前沿和动态,而作为表

征学科领域的科学文献传播网络的演化与发展正是基于先前学科领域研究的基础上，从质的方面得到突破和创新，在不断批判和继承过程中提出符合客观规律的知识。因此，笔者从知识进化这一新的视角对科学文献传播网络进行分析与研究具有可行性，不仅将为科学文献传播网络的趋势发展分析提供新的研究思路，而且也是对知识进化理论研究与应用的深入拓展。

3 科学文献传播网络中的知识进化现象与模式研究

3.1 科学文献传播网络中的知识进化现象及其分析流程

针对知识的主体特征，知识的发生和进化可分为两种基本形式：一是知识的系统进化；二是知识的个体进化。知识的系统进化主要是指具体某一学科的知识作为整体的进化状况。知识的个体进化则指的是各个具体的知识成员如何在人们的认识和实践活动中由初级的观念（知识变异）进化为“完形的”知识形式（知识个体）[11]。科学文献传播网络代表着学科领域一个时期的知识总量，类似于知识的系统进化，也是群体进化，是由各种具体的知识个体构成的（包括表达知识客体内容单元的主题和知识主体外部单元的作者等）。没有知识个体的存在，就不可能有知识群体进化，知识个体的进化是群体进化的前提和基础。而科学文献传播网络是学科领域知识个体的集合，是由知识个体构成的知识群体，也是反映学科领域知识质和量形成及变化的集合。随着文献量的不断增加和文献内容的不断创新，科学文献传播网络在进化动力机制的作用下随时间发生一系列不可逆的进化与演变过程。正是知识进化的一系列进程，引起科学文献传播网络的结构状态发生动态变化，并实现了知识的传承、拓展与创新，最终促进科学的不断发展与社会的不断进步。基于此，笔者构建科学文献传播网络中知识进化现象分析流程图，如图 1 所示：

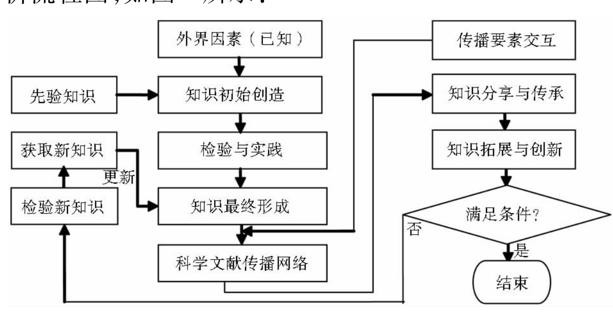


图 1 科学文献传播网络中知识进化现象分析流程

从图 1 中可以看出，首先，知识主体在认识和改造客观世界的过程中，受外界因素的影响，基于自身对客观规律的认识，创造了学科领域知识，并经过检验与实践，选择了那些暂时未被证明是错误的知识；其次，那些被保留的知识以科学文献的方式进行传递与分享，由此构成了以科学文献传播要素承载的知识为内容的科学文献传播网络；再次，科学文献传播网络实现了知识的分享，并在遗传机制的作用下传承与拓展；最终，创造的知识不断被检验，发生知识的变异，促使最终知识符合当前客观规律，同时知识主体也不断调整知识创意，重新创造知识，从而不断产生新的知识个体，以此朝着符合客观规律的方向前进与发展。

3.2 科学文献传播网络中知识进化模式研究

在科学文献传播过程中，传播网络的一系列过程都是遗传与变异共同作用的结果。传播网络的改变并非偶然，是内部因素相互作用同时受外界因素影响，并与之相适应的结果。依据进化机制，知识的变异在环境影响下是具有目的性的，由此表征知识的科学文献传播网络的演变与发展也不是随机的，是具有导向性的[12]。从科学文献传播网络的内容构成，可以分为两部分：一是表征知识客体的传播网络，针对编码化知识进行改进，通常是具有轨道或路径依赖式的进化过程，这类传播网络主要是关键词网络、主题网络；二是表征知识主体的传播网络，更多表现为激进式的知识进化过程，这类传播网络有作者网络、机构网络等。上述两种传播网络都存在着知识变异——选择和保留的过程，且都是具有方向性的，需要经受外界环境因素的选择。其中，主题网络是通过对文献中标题、关键词、摘要乃至全文中主题词的提取与挖掘而构建的网络，对于共现强度弱的主题，将不利于预测潜在的主题或处于上升期的主题，而且在主题的抽取过程中没有统一的标准，存在一定的主观性[13]，在此不考虑采用知识进化理论的方法对其进行演化与预测分析。因此，考虑运用知识进化理论，研究科学文献传播网络的进化过程，以此预测科学文献传播网络的未来发展趋势。

依据知识进化理论中的进化机制：学科领域内知识主体提出的各种知识，符合当前学科价值的则得到更多保留的机会，实现适者生存；在被保留下的知识中，通过科学文献传播方式分享与传递，实现自然选择；在漫长的自然选择过程中，经过不断的知识重组变异积累，最终实现科学文献传播网络的持续进化。进化过程中，科学文献传播网络经历了从无到有、从混沌到清晰、从相似到差异化的过程，并不断调整自身以适

应环境。基于上述理解与分析，笔者描绘了科学文献传播网络中知识进化的模式，如图 2 所示：

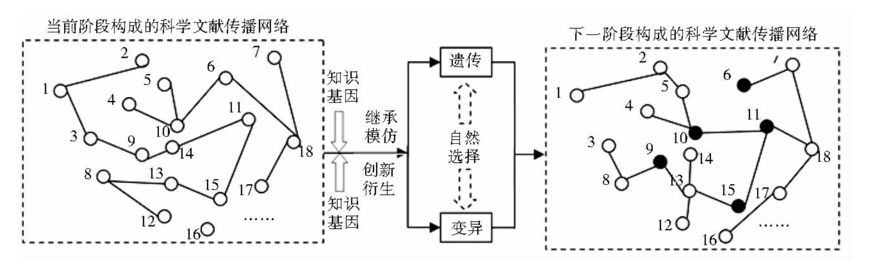


图 2　科学文献传播网络中知识进化模式分析

从图 2 中可以发现，在科学文献传播网络演变进化的过程中，网络结构和节点随之发生动态变化。图 2 中，以 s_m 表示网络结构状态，以 x_i 表示网络单个节点状态。由于受到外界因素的影响，有些网络节点状态发生动态改变，另一些节点则会退出网络。如时间阶段 t 构建的传播网络 $s_m(t)$ 到时间阶段 $t+1$ 构建的传播网络 $s_m(t+1)$ 的进化中，经过自然选择的过程，传播网络 $s_m(t+1)$ 中节点 x_6、x_9、x_{10}、x_{11}、x_{15} 消失了。同时，传播网络中出现了新的节点，在变异机制的推动下，网络节点状态发生了改变，并以遗传的方式保留了下来。另外，从传播网络 $s_m(t)$ 到 $s_m(t+1)$ 的变化过程中，发现网络整体节点的数量、规模等也发生了改变，这也促使科学文献传播网络的知识结构状态会随之发生变化。

从图 2 可以看出，底层结构知识基因是传播网络进化的基础，是组成科学文献传播网络的基本单位和结构要素，也决定着科学文献传播网络的知识进化趋势。科学文献传播网络在进化机制作用下，以知识基因的方式遗传和变异、重组，经过自然选择的过程，呈现螺旋式不断上升的趋势。整个过程不断重现“遗传－变异－选择”三个阶段，并通过知识基因的融合、遗传和拓展、创新等实现的。

由上可知，科学文献传播网络是在人们不断探索客观规律的过程中，在进化机制的作用下随时间而发生一系列的过程。在此过程中，科学文献传播网络的演变与发展主要来自于外界因素的影响，最终在进化机制的驱使下形成了传播网络的演变过程。因此，针对科学文献传播网络未来的发展趋势，可以通过对传播网络的遗传和变异机制，以及进化过程中传播网络的演变路径分析做出判断与预测。基于此，笔者构建科学文献传播网络进化示例图，如图 3 所示。

图 3 的左边是 3 个相关联的科学文献传播网络，即 3 个通过科学文献传播关系形成 3 个网络（见图 3 的中间部分），右图中传播网络上的小箭头表示每个传播网络节点都是有方向的，方向取决于各自的状态属性，这些属性由这一时刻传播网络所包含的内部节点来定义。依据知识进化理论，科学文献传播未来的发展趋势不仅与自身结构有关，更受到来自于外界因素的影响。本文中的外界因素影响体现在与其相关联传播网络对其自身的影响。由此，科学文献传播网络之间的相互交叉引用，充分体现

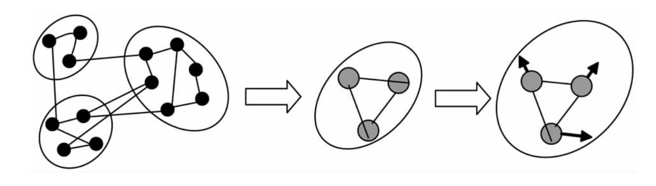


图 3　科学文献传播网络进化实例

了它们之间的关系紧密度，同时关联性的强弱又决定了传播网络之间的相互影响程度。由此，科学文献传播网络未来的发展趋势取决于与之相关联网络结构的特性和它们之间联系程度的强弱，最终在进化动力机制的推动下实现前进与发展的。

4　知识进化视角下科学文献传播网络趋势分析与研究

本节依据知识进化理论的进化机制，阐述推动科学文献传播网络前进的驱动力，结合科学文献传播网络演变过程，判断传播网络驱动力的区域空间，由此构建基于进化视角的科学文献传播网络预测模型，为科学文献传播网络的发展趋势提供参考和借鉴。

4.1　知识进化视角下科学文献传播网络演化机制分析

依据知识进化理论，科学文献传播网络进化过程中始终存在着两种作用力：一种是遗传作用力，主要是促进传播网络继承、衍生的动力；另一种是变异作用力，主要是促进传播网络拓展、创新的动力。两种作用力相互作用、相互制约，共同推动着科学文献传播网络的发展。针对科学文献传播网络进化机制的研究，笔者借助达尔文的进化论思想，用 f_m 表示遗传作用力，f_n 表示变异作用力，N_m、N_n 分别为遗传作用力和变异作用力的要素的集合，则科学文献传播网络进化过程中遗传作用力和变异作用力分别表示为：

$$f_m = \frac{\partial N_m}{\partial t};f_n = \frac{\partial N_n}{\partial t} \qquad 式(1)$$

对式(1)进行分析，可以得到以下 4 种结果：

(1)若 $f_m > f_n$,网络的驱动力主要以遗传、衍生为主。此时,网络结构形态得以延续,科学文献传播网络更多体现的是继承关系,整个学科领域也不会有较大的改变。

(2)若 $f_m < f_n$,网络的驱动力主要来自拓展、变异与创新方面。此时,由于外界环境因素的影响,导致整个网络结构受到较大的影响,从而促使科学文献传播网络形态发生改变,呈现进化的发展态势。

(3)若 $f_m \neq f_n$,网络的驱动力来自遗传、衍生和拓展、变异两个方面。此情况很可能出现在传播网络进化过程的成长期。当前阶段,学科领域内的研究者由于外界因素的影响,对学科领域知识的认知度有了较大的改变,导致科学文献呈现的知识主题、作者和机构也发生了改变,引起网络形态发生迁移,网络呈现进化的发展态势。该阶段是科学文献传播过程的关键时期,既有知识的传承,也有知识的拓展,最终为传播网络的进化发展奠定了基础。

(4)若 $f_m = f_n = 0$,网络的两种作用力都趋向于零,说明网络没有了动力来源,结构会逐渐解体,这可能出现在进化的衰退期。在此期间,研究人员对于学科领域内的发展前景不太看好,学科领域的发展已遇到瓶颈,网络结构会逐渐退化,直到解体为止。

科学文献传播网络的进化过程,正是由于外界因素的影响,在遗传作用力与变异作用力的推动下,促使传播网络结构逐渐呈现出无序、不均衡的状态,一旦量变累积到一定程度时,将打破当前固有的网络稳定状态,从而使科学文献传播网络系统结构跃迁到一个更高层次。一般情况下,内外环境发生剧烈变化时,知识的量变累积到一定程度必然引起质变,促使科学文献知识有了"质"的突破,由此导致科学文献传播网络处于质变跃迁的过程。

4.2 知识进化视角下科学文献传播网络预测模型

科学文献传播网络发展趋势的判定基本思想是:传播网络未来的变化值,一方面遗传于传播网络自身的值的大小,另一方面由于变异而产生的新的值,由两方面共同组合而产生的。通过传播网络不同时期变异强度的大小可以判断,传播网络遗传的值和变异的值之间的比例关系,从而预测科学文献传播网络未来的趋势。根据张华夏总结的卡尔·波普尔证伪主义和进化认识论的思想[14],笔者以遗传作用力和变异作用力为动力机制,构建科学文献传播网络预测值的计算公式,如式(2)所示:

$$K_{t+1} = \{(K_t, P_t), (W_t, Q_t), (N_{t+1})\} \quad 式(2)$$

在式(2)中,K_{t+1} 是在 $t+1$ 时间阶段传播网络的预测值,K_t 是传播网络在 t 时间段的值,W_t 是传播网络在 t 时间段由变异产生的值,N_{t+1} 表示传播网络在 $t+1$ 时间段 K_{t+1} 所取预测值的个数。P_t、Q_t 分别是传播网络在 t 时间段继承关系与新生关系的权重系数,两者的取值变化在 0 到 1 之间。针对 P_t、Q_t 值的大小,参考由张建华提出的知识进化绩效测度指标评分策略和权重配置方案[15],结合科学文献传播网络的生命周期特征,判断引起传播网络演变与发展的遗传作用力和变异作用力的强度差异,进而确定权重系数 P_t、Q_t 的具体值,具体如下:

(1)萌芽期。考虑此阶段变异强度较小,表现出的变异作用力较弱,由此传播网络的继承关系占主导地位,遗传权重系数较大,P_t、Q_t 可以分别取值为 0.7、0.3;

(2)成长期。考虑此阶段变异强度较大,表现出的变异作用力较强,由此传播网络的新生关系占主导地位,变异权重系数较大,P_t、Q_t 可以分别取值为 0.3、0.7;

(3)成熟期。考虑此阶段变异强度一般,表现出的变异作用力也一般,由此传播网络的继承关系与新生关系同等重要,遗传和变异权重系数比较平衡,P_t、Q_t 可以分别取值为 0.5、0.5;

(4)衰退期。此阶段遗传作用力与变异作用力趋于零,P_t、Q_t 近似取值为 0。

依据知识进化理论,科学文献传播网络结构状态未来的趋势来自于外界因素的影响,主要体现在与其相关联传播网络结构的影响程度。笔者假设学科领域内构建了科学文献传播网络 i,在时间阶段 t 的结构状态标记为 s_m,当忽略外界因素的影响时,进化过程中 s_m 的结构状态将以遗传的方式传递,科学文献传播网络的状态不会发生改变,则 P_t 为 1,Q_t 为 0,所以 s_m 在时间阶段 $t+1$ 的发展趋势为:

$$s_m(t+1) = f[s_m(t)] \quad 式(3)$$

式(3)是在没有外界因素的影响下科学文献传播网络未来的趋势发展是由当前传播网络的状态所决定的,其中 $s_m(t+1)$ 表示传播网络在 $t+1$ 时刻的网络结构状态。如果科学文献传播网络 s_m 在当前阶段受到传播网络 s_{n-1} 上一阶段 $t-1$ 的影响,即它们之间的相互作用强度可表示成 $l_{m(t)n(t-1)}$。以 $x_1, x_2, \cdots, x_i$ 表示科学文献传播网络节点,$x_n(t)$ 表示传播网络节点在 时刻的特征属性。由于外界因素的影响,知识主体对客观规律的认知也发生改变,进一步促进了引起科学文献传播网络的变异与创新。因此,为完善科学文献传播网络的趋势变化情况,对式(3)做了必要的修改,具体

如式(4)所示：

$$s_m(t+1)=P_t * f[s_m(t)]+Q_t * \sum_{m=1,n=1}^{i} l_{m(t)n(t-1)} x_n(t) \quad m,n \in [1,2,\cdots,i] \qquad \text{式(4)}$$

式(4)右边的第一部分$f[s_m(t)]$表示s_m在时间阶段$t+1$时的发展趋势对自身阶段t的状态继承关系，这主要是来源于人们常常使用知识的当前状态表示未来发展的趋势，因此这里直接取线性函数$f[s_m(t)]=s_m(t)$。等式右边的第二部分则表示s_m在时间阶段$t+1$时的发展趋势对相关联传播网络在时间阶段t时的依赖关系，$l_{m(t)n(t+1)}$表示传播网络$s_m(t)$与$s_n(t-1)$之间的连接强度[16]，这里取

$$l_{m(t)n(t-1)}=h_{m(t)n(t-1)}/\sum_n h_{m(t)n(t-1)} \qquad \text{式(5)}$$

在式(5)中，$h_{m(t)n(t-1)}$为$s_m(t)$的节点数与$s_n(t-1)$的节点数两者连接数之和[17]。由于科学文献传播网络之间的关联对整个科学文献传播网络的趋势发展具有重要影响，而关联系数采用的是网络节点之间的连接强度，连接强度在每一时间阶段都不一定完全相同，由此形成的一种非线性关系的时变网络也是随时间发生动态变化的。

综上所述，科学文献传播网络的发展趋势受到自身及与其相关联传播网络的影响，影响程度与传播网络之间的连接强度成正比。依据知识进化理论，当忽略外界的影响因素时，传播网络在进化过程中网络结构状态不会发生改变，将保持原有的状态。但是在外在因素的影响下，传播网络之间的相互影响，促使网络结构状态也随之发生动态变化。

5 实证分析——以科学文献关键词网络为例

由于不同学科的各自特性存在差异，由此呈现的演化规律也有所差别，笔者仅以传播网络节点之间的关系与形成的不同结构为研究对象，分析知识主体的外部单元与知识客体的内容单元的相互作用及其形成的不同关联，探究科学文献传播要素之间复杂演化的特点和规律。本文以科学文献关键词为网络例进行实证分析，科学文献中的关键词是表征文献主题内容具有实质意义的词语，是对揭示和描述文献主题内容来说是重要的、带有关键性的那些语词，可以用于描述一篇文献的主要特征，代表着文献主体的研究动态和文献主题的分布领域。同时，根据关键词构成中的词汇分布情况以及关键词的共现，关键词之间存在一定的连通性，一定规模的关键词会构成一个有效的网络，形成关键词网络。关键词网络体现的共现关系体现了文献核心内容的亲疏关系和紧密关系[18]。因此，可以通过对文献关键词共现的网络特征分析，以挖掘科学文献内容之间的衔接关系和文献主体之间研究的交叉点。由此，笔者以新能源领域的数据为统计源，运用知识进化视角下的演化与预测方法，对1980-2016年间新能源子领域风能的关键词网络进行演化分析，并按照预测模型测算出2017年的数据，将预测的结果与实际的数据进行对比分析，以此验证预测模型的有效性和可行性。

5.1 数据处理

数据来源于CNKI，选取中国学术期刊网络出版总库下的工程科技Ⅱ辑的新能源类目，按时间排序获取相关数据。统计新能源领域1980-2017年期间发表的所有文献的特征项，统计时间截止到2017年9月底。选择的新能源领域成长周期相对较长，样本时间横跨1980-2017年，从数据处理的角度出发，由于1980-1983年文献量较少，所以将1980-1983年的文献量进行合并，作为初始年份时间窗口，将2017年作为截止年份时间窗口，如图4所示。图5显示了新能源关键词共现频次大于16的网络图。

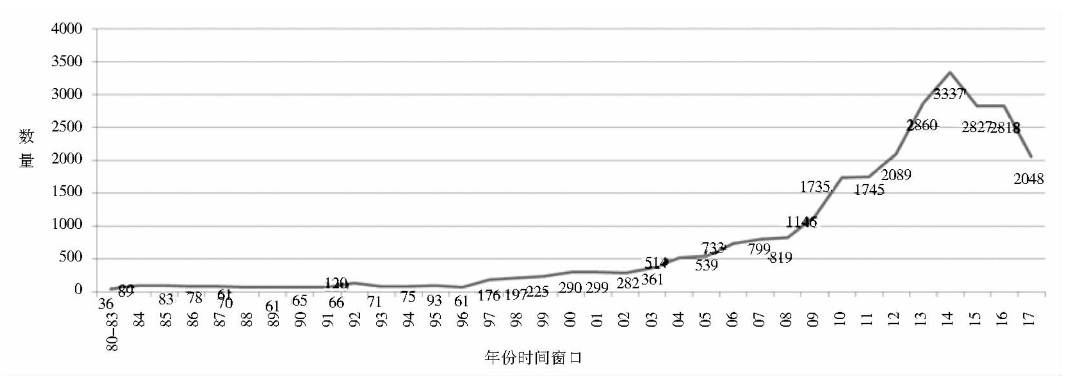


图4 时间片划分分布

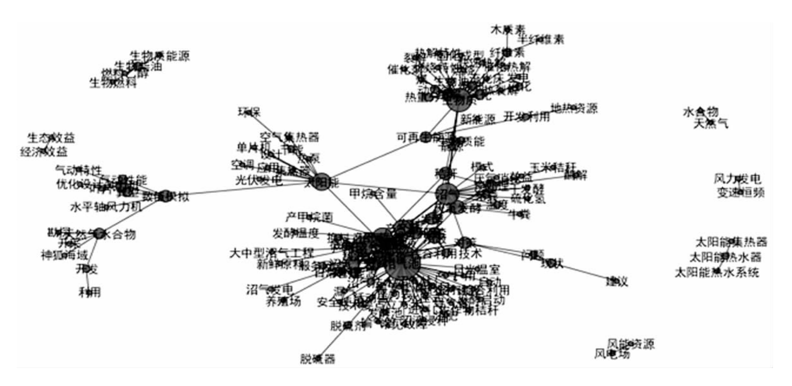


图 5　新能源关键词共现网络图(共现频次大于 16)

由图 4 可以看出,从 1980 年开始出现新能源的相关研究,文献增长量在相当长一段时间内比较小,也没有发生较大的波动,文献量维持在 100 篇左右,直到 1997 年文献量出现了一个较快的增长。而到了 2009 年,新能源领域的文献量增长到 1 000 篇以上,开始呈现出较快的指数增长趋势。综合以上分析,笔者将中国新能源领域科学文献传播网络按照生命周期理论划分为三个阶段:

(1)孕育阶段:1980 – 1996 年。这一阶段该领域的年度文献增长率在 0 – 0.2 区间范围。符合生命周期理论中的孕育期特点,该阶段新能源领域刚刚起步,文献数量较少且文献增长也不明显。

(2)成长阶段:1997 – 2008 年。该领域在这一阶段的年度文献增长率在 0.12 – 0.4 区间范围内。该阶段文献增长率震荡不稳定,有先递增后稍有减少再增长的趋势,但是文献增长量相比较孕育阶段要大的多。

(3)成熟阶段:2009 – 2017 年。这一阶段的年度文献增长率在 0.3 – 0.6 区间范围。文献增率呈较为陡峭的指数增长式,无论是文献数量还是文献增长率都较之前的阶段有极大幅度的增长。

5.2　基于风能的关键词网络演化

为了印证笔者提出的知识进化视角下的科学文献传播网络预测模型的可行性和有效性,本节选取新能源领域中的风能这个研究主题作为案例进行分析。之所以选择风能主题,有两方面的原因:①由于新能源领域文献量较大,分析起来有一定的难度,所以有必要选择一个子领域进行实证分析;②风能的研究过程贯穿了所有时间窗口,数据完整,具有较好的代表性。通过设置词频阈值,得到风能的关键词网络演化分布,如表 1 所示:

表 1　基于风能的关键词网络演化分布

时间窗口	主要关键词	词频统计(阈值)
时间窗口 2007 年	风力发电、风电场、风能密度、风能资源、风力发电机、动力学、表面气流、表面过程、风能密度	5
时间窗口 2008 年	风力发电、风力发电机、风电场、风能密度、风能资源、风能发电、风能提水、风力机、电力系统	6
时间窗口 2009 年	风力发电、风电场、风能发电、风力致热、风能资源、能量转换、风能利用、可再生能源、风力发电机	5
时间窗口 2010 年	浓缩风能型风力发电机、风力发电、风电场、风能密度、迎风、最大风能捕获、风能资源、风能转换、容量系数、风能发电	6
时间窗口 2011 年	最大风能捕获、控制系统、风能转换、风力发电机、风电场、风能发电、风能密度、风电场存储模式、风力发电、风能资源	6
时间窗口 2012 年	最大风能捕获、风力发电浓缩风能型风力发电机、风电场、控制系统、新能源、风能密度、变转速风力发电系统、风力机、风能资源、风力发电机	6
时间窗口 2013 年	风力发电、浓缩风能型风力发电机、风能密度、最大风能捕获、控制系统、变转速风力发电系统、新能源、风力发电机、风力机、变速恒频风力发电系统	6
时间窗口 2014 年	浓缩风能型风力发电机、风力发电、矢量变换控制、风力发电机、风能密度、数值模拟、变速恒频风力发电系统、控制系统、单片机、风能资源、最大风能利用搜索算法、变转速风力发电系统、相似模型、同步发电机	7
时间窗口 2015 年	风力发电、浓缩风能型风力发电机、最大风能捕获、变速恒频、风能密度、变速恒频风力发电系统、风能转换系统、数值模拟、功率分流、风能利用、风能资源、变转速风力发电系统	8
时间窗口 2016 年	风能转换系统、浓缩风能型风力发电机、风力发电、最大风能捕获、变速恒频、动态补偿、流场特性、相似模型、开关磁阻发电机、风能资源、风能资源评估、风能密度、最大风能跟踪	8
时间窗口 2017 年	风能转换系统、浓缩风能型风力发电机、变速恒频、最大风能跟踪、变转速风力发电系统、开关磁阻发电机、流场特性	9

从表 1 中可以看出最近 10 年间,风能领域的关键词网络演化过程,在 2007 – 2009 年间,风电场的建设一直是研究的热点,而随着研究的深入,最大风能捕获和浓缩风能型风力发电机逐渐取代了风电场成为研究热点。从中可以发现知识是不断进化的,在遗传和变异这两种机制下,风能领域的研究主题不断演化,但这种演化是循序渐进的,而不是突变的。

同时,从图 4 中可以发现,新能源年度文献增长率在 0.3 – 0.6 区间范围,处于成熟期阶段,此阶段的特征是:变异强度一般,表现出的变异作用力也一般,由此得知风能等各个子领域构建的关键词网络表现出的继承关系与变异关系同等重要。因此,风能的关键词网络遗传权重系数和变异权重系数 P_t、Q_t 可以分别取值为 0.5、0.5。

5.3　基于风能的关键词网络预测分析

新能源领域中主要包括风能、太阳能、生物质能、

地热能、海洋能、化学能、原子能等子领域，上述子领域之间的节点联系紧密，内部连接也很紧密。通过新能源领域数据采集与分析，构建风能、太阳能、生物质能、地热能、海洋能、化学能和原子能等子领域的关键词网络，由此得知新能源领域 的组成内容，如下：

$$s = \{s_{fn}, s_{tyn}, s_{swzn}, s_{drn}, s_{hyn}, s_{hxn}, s_{yzn}\} \quad 式(6)$$

上述式(6)右边表示风能、太阳能、生物质能、地热能、海洋能、化学能和原子能等子领域的关键词网络，与风能构建的关键词网络之间的连接，可以求得预测模型中 $h_{m(t)n(t-1)}$ 的大小，进而计算出风能构建的关键词网络与太阳能等子领域构建的关键词网络之间的连接强度 l_{fn}，结果如式(7)所示。

$$l_{fn} = \{l_{fn\text{-}tyn}, l_{fn\text{-}swzn}, l_{fn\text{-}drn}, l_{fn\text{-}hxn}, l_{fn\text{-}yzn}\} = \{0.41, 0.17, 0.15, 0.13, 0.09, 0.05\} \quad 式(7)$$

式(7)显示了风能与各个子领域连接强度的值，如风能与太阳能、海洋能的连接强度为0.41、0.13。$x_1, x_2, \cdots, x_i$ 表示不同的关键词网络节点，$x_n(t)$ 表示不同领域构建的关键词网络在 t 时刻的特征属性。W_t 是关键词网络在 t 时间段受变异作用产生的值，通过各个子网络的当前值和表2显示的网络之间连接强度可以得知。

$$W_t = \sum_{m=1,n=1}^{i} l_{m(t)n(t-1)} x_n(t) \quad m, n \in [1, 2, \cdots, i] \quad 式(8)$$

通过式(8)可以求得基于变异作用力所获得的风能关键词集合 W_t。K_t 是基于风能的关键词网络在2016年时间段的值，通过采集的数据已经得知。同时，依据表1关于风能的关键词网络分布特征，为了清晰地显示风能的关键词网络在2017年主要的关键词，笔者设置网络在 $t+1$ 时间段所取预测值 N_{t+1} 为20。结合之前的遗传权重系数和变异权重系数 P_t、Q_t 分别为0.5、0.5。由此对风能领域关键词网络进行预测，如式(9)所示：

$$K_u = \{(K_t, 0.5), (W_t, 0.5), 20\} \quad 式(9)$$

通过上述内容分析，得知基于风能的预测值主要有两部分组成：一部分来自遗传作用力获得的遗传值，表2中 K_t 所示，以及遗传的权重系数 P_t 的值为0.5，由此求得风能关键词网络依靠继承关系所获得的值；另一部分来自变异作用力获得的变异值，表2中 W_t 所示，以及变异的权重系数 Q_t 的值为0.5，从而求得风能关键词网络依靠变异关系所获得的值。最终，依据网络预测模型式(9)，得出风能在 $t+1$ 时间段的预测值 K_u，数据结果见表2。

在表2中，对风能领域关键词网络中的关键词词频预测值按照词频进行排名，推荐了前20个关键词。

表2　基于风能的关键词网络数据

数据类型	主要关键词
遗传值 K_t	风能转换系统、风力发电、最大风能捕获、风能、变速恒频、风能资源评估、风能密度、最大风能跟踪、发电机、浓缩风能型风力发电机、流场特性、相似模型、开关磁阻发电机、新能源、电力、风能资源、风能发电系统、风电场、风能资源、动态补偿
变异值 W_t	风电、产气性能、新能源、能量储存、风能资源、滑膜控制、可再生资源、测控系统、资源评估、模糊控制、动态补偿、数值模拟、动态特性、热性能、能量效率、利用系数、控制性能、太阳辐射模型、发电、追踪系数
2017 真实数据	风能转换系统、浓缩风能型风力发电机、变速恒频、最大风能跟踪、变转速风力发电系统、开关磁阻发电机、流场特性、模糊控制、最大风能捕获、风能转换、控制系统、可再生资源、风力发电系统、新能源、测控系统、风能资源、风力机、能量存储、滑膜控制、资源评估
2017 预测数据	风能转换系统、变转速风力发电系统、风力发电、最大风能捕获、风能、变速恒频、风能资源评估、风能密度、最大风能跟踪、风能资源、发电机、浓缩风能型风力发电机、风电、风力机、新能源、能量存储、滑膜控制、流场特性、模糊控制、测控系统

从表2中对风能关键词网络在2017年真实数据与预测数据对比分析，发现共有14个关键词是相同的，准确率为70%。同时，从表1和表2中可以发现，预测的关键词网络既遗传了以前关键词网络的特征，也通过变异机制，有了新的趋势。风能转换系统、变转速风力发电系统和最大风能捕获一直都是近期的研究热点，在预测中的排名也是比较靠前，是知识遗传作用力的结果。而滑膜控制、模糊控制、测控系统等系统控制方法的兴起则是由风能转换系统异变出来的，体现了知识变异作用力

该实例通过模型预测2017年风能关键词网络的排名靠前的数据，通过数据结果对比分析验证了笔者提出的知识进化视角下科学文献传播网络趋势预测方法的可行性和有效性，能够根据历史数据对未来时间段关键词网络进行预测和判断。

6　结语

笔者针对从海量科学文献中难以准确而快速地梳理相关领域研究发展动态的问题，提出从科学文献传播过程中由各个传播要素构建的科学文献传播网络的关联分析为切入点，以知识进化理论为视角，剖析科学文献传播网络的演化机制，构建科学文献传播网络预测模型，以此探索科学文献传播未来的动态发展趋势，为揭示科研新成果、挖掘学科领域研究热点提供研究思路和参考借鉴。

目前，还鲜有从知识进化视角判断与预测科学文献传播网络的研究，这将为科学文献传播网络的趋势发展分析提供新的研究思路，也是对知识进化理论研

究与应用的深入拓展。但是,如果单纯依赖知识进化理论中的进化动力机制还很难挖掘出蕴含在科学文献中的动态信息,为了能够更加准确地从大量的科学文献中预测出学科领域未来的研究主题和动态的发展趋势,下一步工作将是从知识进化中的量变和质变两个方面,结合复杂网络分析方法,对科学文献传播网络进行更加深入的探索,以期为科技创新和科技决策提供支持和帮助。

参考文献:

[1] 邹常诗. 科学文献计量分析与文献关联性研究[J]. 情报资料工作,2000(4):18 – 20.
[2] 王燕鹏. 国内基于主题模型的科技文献主题发现及演化研究进展[J]. 图书情报工作, 2016,60(3):130 – 136.
[3] 达尔文. 物种起源[M]. 翟飚,译. 北京: 人民日报出版社, 2005.
[4] 波普尔. 科学知识进化论[M]. 纪树立,译. 北京: 三联书店, 1987.
[5] TANRIVERDI I. Diffusion of telemedicine: a knowledge barrier perspective[J]. Telemedicine journal, 2004,5(3):223 – 244.
[6] 吴力群. 知识基因、知识进化与知识服务[J]. 现代情报, 2005 (6):177 – 179.
[7] 刘植惠. 知识基因理论新进展[J]. 情报科学, 2003(12): 1243 – 1245.
[8] 张凌志, 和金生. 基于生物进化模式下的知识进化机理研究[J]. 情报科学, 2011,30(2): 105 – 109.
[9] INKPEN A C, TSANGE W K. Social capital networks and knowledge transfer[J]. Academy of management review, 2005,30(1): 146 – 165.
[10] 王曰芬, 丁玉飞, 刘卫江. 基于知识进化视角的科学文献传播网络演变研究[J]. 情报资料工作, 2016(2):5 – 10.
[11] MARCO T, LESLIE L. Empiricalanalysis of the evolution of a scientific collaboration network[J]. Statistical mechanics and its applications,2007,383(2):750 – 764.
[12] 赵健宇, 李柏洲. 对企业知识创造类生物现象及知识基因论的再思考[J]. 科学学与科学技术管理, 2014,35(8):18 – 27.
[13] 唐果媛,张薇. 基于共词分析法的学科主题演化研究进展与分析[J]. 图书情报工作,2015,59(5):128 – 136.
[14] 张华夏. 波普尔的证伪主义和进化认识论[J]. 自然辩证法研究, 2003(19): 10 – 13.
[15] 张建华. 知识管理中的知识进化绩效评价机制研究[J]. 科学学与科学技术管理,2013, 34(7): 28 – 37.
[16] 刘向. 知识网络的形成与演化[M] . 北京: 武汉大学出版社, 2014.
[17] BOCCALETTI S, LATORA V, CHAVEZ M, et al. Complex networks: structure and dynamics[J]. Physics reports, 2006,424(4 – 5): 175 – 308.
[18] 吴晓秋, 吕娜. 基于关键词共现频率的热点分析方法研究[J]. 情报理论与实践, 2012, 35(8): 115 – 119.

作者贡献说明:

丁玉飞:提出论文研究思路,负责数据收集与实证分析,并撰写和修改论文;
关鹏:审阅论文,提出论文修改建议。

Study and Application of Translation and Prediction of the Scientific Literature Communication Network from the Perspective of Knowledge Evolution

Ding Yufei Guan Peng
School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094

Abstract: [**Purpose/significance**] Analyzing and researchingthe development trendof scientific literature communication network can provide reference to the following work: reveal new scientific research achievements, grasp the development context of science and technology context and find the academic foreground. [**Method/process**] At first, this paper expounded the theory of knowledge evolution and its application. And then, it discussed the feasibility of the scientific literature communication network forecasting method from the perspective of knowledge evolution. Finally, it constructed the forecasting model of scientific literature communication network by analyzing the evolution mechanism of science literature communication network, and verifiedthe forecasting modelthough the keywords network of scientific literature. [**Result/conclusion**] The empirical analysis of scientific literature shows that the evolution and trend forecasting method of scientific literature communication network is feasible and effective. Besides,this model can providesnew research ideas for the trend analysis of science literature communication network.

Keywords: knowledge evolution scientific literature communication network translation prediction